

## REMARKS

Attached hereto is a marked-up version of the changes made to the specification and the claims by the current amendment. The attached Appendix is captioned **"Version with markings to show changes made."**

The specification has been amended to remove minor grammatical errors and to place the application in better form. No new matter has been added. New claims have been added and existing claims have been amended to more clearly describe the present invention.

Consideration and allowance of the subject application is respectfully requested.

Respectfully submitted,

GREER, BURNS & CRAIN, LTD.

By



B. Joe Kim

Reg. No. 41,897

July 15, 2002

300 S. Wacker Drive – Suite 2500

Chicago, Illinois 60606

Tel.: (312) 360-0080

Fax: (312) 360-9315

Customer No.: 24978

K:\0321\65712\PRE-AMD.doc

1201.66170



COPY OF PAPERS  
ORIGINALLY FILED

RECEIVED

JUL 25 2002

TECH CENTER 2800  
Serial No. 10/087,226

**VERSION WITH MARKINGS TO SHOW CHANGES MADE**

**In the Specification:**

The paragraph beginning on 4, line 9 has been amended as follows:

Turning now to FIG. 2, the present invention relates to a model of the superior colliculus (SC) 10 of a vertebrate brain 12 (shown in FIG. 1), which integrates multisensory input and guides orienting movements. The model 13, as in the SC 10 of the brain 12, are organized as a map 14 having a plurality of grids or units 16. Each unit 16 on the map 14 represents a collicular neuron that receives multisensory input from its corresponding location in the environment. The units 16, i.e., the model SC units 16 use sensory inputs such as video (V) 18 and audio (A) 20, for example, to compute the probability that something of interest, i.e., a target 22, has appeared in the surroundings.

The paragraph beginning on 4, line 17 has been amended as follows:

~~A—~~The model 13 in accordance with one embodiment of the present invention approximates Bayes' rule for computing the probability of a target. Specifically, the SC units 16 in the map 14 approximate  $P(T|V,A)$ , which is the conditional probability of a target (T) given visual (V) and auditory (A) sensory input. The Bayes' rule for computing the probability of a target given V and A is as follows:

The paragraph beginning on page 6, line 3, has been amended as follows:

~~A~~—The model 13 in accordance with another embodiment of the present invention estimates Bayes' rule for calculating target probability by using back-propagation which, as known in the art, is a supervised neural network learning algorithm. Generally, back-propagation is used to train neural networks having input units, output units, and units in between called hidden units. All units are sigmoidal. The input units send their activity to the hidden units, and the hidden units send their activity to the output units. The hidden and output units can also receive a bias input, which is an input that has activity 1 all the time. All the connections between the input, output, and hidden units have weights associated with them. Back-propagation adjusts the values of the weights in order to achieve the desired output unit response for any input pattern. In the estimation method, the SC units 16 are the output units of neural networks that also have input and hidden units. The back-propagation algorithm is used to iteratively adjust the weights of the hidden and the output units to achieve the desired output.

The paragraph beginning on page 6, line 26, has been amended as follows:

Referring to FIG. 6, ~~The~~ the acquisition phase includes acquiring raw video and audio input and preprocessing it (block 54), and applying the input to the neural network and finding the responses of the SC units 16 (block 56). Then, the SC unit 16 with the highest value is found (block 58). Using this information, the location

corresponding to the SC unit 16 with the highest response value is chosen as the location of the next target (block 60).

The paragraph beginning on page 7, line 11, has been amended as follows:

Turning now to FIG. 7, the present unsupervised algorithm for approximating target probability includes two stages. The first stage involves an unsupervised learning mechanism that increases the amount of information transmitted from the sensory inputs, audio (A) and video (V), for example, to the SC unit 16 of the model SC\_13. This mechanism is known in the art as the Kohonen mechanism, which has been shown to increase information transmission in neural networks. The Kohonen mechanism is unsupervised, meaning that it would take the sensory inputs (such as audio and video) and automatically adjust the model SC\_13 to increase the amount of information that is transmitted to it from the input. This is accomplished by adjusting the connection weights from the V and A inputs to the SC units 16 in such a way that individual SC units become specialized for specific inputs. For example, the Kohonen algorithm might cause one SC unit 16 to become specialized for video input from the extreme left side of the environment, and another to become specialized for audio input coming straight ahead. For very certain (not noisy) inputs, all the SC units 16 will become specialized for particular locations in the environment, and almost all of them will become specific for one modality or the other (V or A). The SC units 16 in this case

can give a near maximal amount of information about the input. These units 16 can indicate not only where the target is but also of what modality it is.

The paragraph beginning on page 7, line 29, has been amended as follows:

If the inputs are not so certain (noisy), then the Kohonen algorithm will cause more of the SC units 16 to become bimodal and respond to both V and A. These SC units 16 would be less informative because they could indicate where the target is but not of which modality it is. Thus, the Kohonen algorithm will do the best it can with the input it is given to increase the amount of information that is transmitted to the SC units 16 from the V and A input units.

The paragraph beginning on page 8, line 20, has been amended as follows:

Cortical units 62 modulate the sensory inputs to the model SC units 16 by multiplying their weights. For example, the video input to an SC unit 16 would be  $c_v w_v V$ , where  $c_v$  is the amount of cortical modulation of that sensory weight  $w_v$ . In the learning process, an active cortical unit 62 will increase its modulation of a sensory input to an SC unit 16 if the SC unit is also active but the sensory input is inactive. If the SC unit 16 and the sensory input are both active then the cortical unit 62 will decrease its modulation of the sensory input. For example, when an SC unit 16 receives multisensory video and audio input after stage one training, and a target appears that provides a video input but produces no audio input, that SC unit will be active because it receives both

video and audio input and the video input is active. A cortical unit 62 sensitive to video will also be active. Because the activity of the SC unit 16 and the cortical unit 62 are correlated, the cortical unit will change its level of modulation of the sensory inputs, accordingly as they are anti-correlated. Specifically, the cortical unit 62 will decrease its modulation of the video input (because the cortical unit and the video input are correlated) but increase its modulation of the auditory input (because the cortical unit and the audio input are anti-correlated).

The paragraph beginning on page 9, line 6, has been amended as follows:

Turning now to FIG. 8, the preferred embodiment for implementing the two-stage algorithm for approximating target probability involves iterative procedures that begin after certain parameters in the model have been set. The structure of the neural network model 13 is determined in block 64, in which the number of SC unit 16 is set, and the bias weight and sensitivity of each SC unit are assigned. All the SC units 16 in the two-stage model are sigmoidal, where output  $y$  is related to input  $x$  by:  $y=1/(1+\exp(-gx))$ . The input  $x$  is the weighted sum of its inputs from  $V$  and  $A$  and from the bias. The bias weight  $w_b$  is the same fixed constant for all SC units 16. The sensitivity  $g$  is another fixed constant that is the same for all SC units 16. These fixed constants ( $w_b$  and  $g$ ), along with the number of SC units 16, are set in block 64.

The paragraph beginning on page 10, line 12, has been amended as follows:

Referring now to FIG. 9, each iteration of the stage one learning process begins by acquiring and preprocessing the video and audio inputs from a randomly positioned target (block 76). These V and A inputs are sent to the SC units 16 over the ascending connections. As explained above, the sigmoidal SC units 16 use the weighted sum of these inputs to compute their responses (block 78). Then the SC unit 16 with the maximal response is found (block 80). The unit with the maximal response is referred to as the 'winning' SC unit. The ascending weights of the winning SC unit 16 and its neighbors are trained using Kohonen's rule (block 82). The neighbors of an SC unit 16 are simply the other SC units that are near it in the network. The number of neighbors trained in stage one is determined by the neighborhood-size parameter set in block 77 (see FIG. 8). Kohonen's rule basically adjusts the ascending weights to the winning SC unit 16 and its neighbors so that they become even more specialized for the current input.

The paragraph beginning on page 10, line 24, has been amended as follows:

Turning to FIG. 10, each iteration of stage-two learning process begins by acquiring and preprocessing the video and audio inputs from a randomly positioned target, and using that input to determine cortical activation (block 84). The term 'cortical' is meant to indicate that these units 62 are at a high level, as they are in the cortex of the mammalian brain, and the properties of the cortical units 62 can vary over a very broad range. For example, the cortical units can act as pattern recognizers, and can be specialized for particular types of targets like humans or airplanes. So far as applied here, the cortical units 62 simply register the modality of the target, whether it is visual, auditory, or both.

A visual cortical unit 62, for example, would be active whenever the video input is active. Block 84 indicates that the activity of the cortical units 62 is dependent upon the video and audio inputs. The cortical units 62 send descending connections to the model SC units 16, and more specifically, to the connections onto the SC units from the V and A sensory inputs. As explained above, an active cortical unit 62 can modulate the weights of the ascending connections by multiplying the value of the ascending weight by that of the descending weight (block 86). After any cortical descending modulation of ascending weights is taken ~~input~~ into account, the responses of the SC units 16 to the ascending input is computed (block 88).

The paragraph beginning on page 11, line 11, has been amended as follows:

Then the SC units 16 with responses less than cutoff are found and set to zero (block 90). Descending weights of SC units 16 are then trained using the following triple correlation rule (block 92):

If an SC unit 16 and a cortical unit ~~62~~ are 62 are both active, then  
*increase* the descending weights to *inactive* ascending input synapses, and  
*decrease* the descending weights to *active* ascending input synapses.

The paragraph beginning on page 13, line 5, has been amended as follows:

The ~~Selection~~ selection process is implemented by choosing the SC unit 16 that has the largest response to its inputs. Since the SC units 16 are in spatial register with their inputs, localization of the target is determined by the location of the chosen SC unit in the 1-dimensional array. Acquisition of the target then takes place by moving the rotating platform 102 to the coordinate in the environment corresponding to the chosen



SC unit, thereby allowing the target 100 to be viewed by the operator through a monitor  
116.

The paragraph beginning on page 13, line 11, has been amended as follows:

If target probability is obtained by estimating Bayes' rule using back-propagation, an array of computer-controlled buzzer/flasher pairs (not shown), spaced every 15 degrees, for example, is used to provide the sensory stimuli for back-propagation training. ~~At~~ At each training cycle, one location is chosen at random, and the buzzer and the flasher at that location are activated. The 60 preprocessed audio and video signals are temporally summed or averaged over a window of 1 second and applied as input to the model. The inputs are applied to the model SC units 16 directly, or indirectly through a network of hidden units. The location of the source is specified as a 60 element desired output vector of 59 zeros, and a one at the location in the vector corresponding to the location of the source. The weights are all trained with one cycle of back-propagation, and the process is repeated with a source at a newly chosen, random location.

The paragraph beginning on page 13, line 22, has been amended as follows:

After training, the inputs are preprocessed as described above over the 1-second window and then applied in spatial register to the SC model 13. Each SC unit 16 then estimates, on the basis of its video and audio inputs, the Bayesian probability that the source is present at its corresponding location in the environment, which simulates MSE. The location of the model SC unit 16 with the largest response is then chosen as the location of the most probable target, and the camera 96 and the microphone 98 are

aimed in that direction. The SAC system 94 chooses as targets those objects in the environment that move and make noise, which covers most of the targets actually chosen by the SC in guiding saccadic eye movements in animals.

In the Claims:

Claims 5, 9-12 and 20-24 have been canceled; claims 1, 6-8 and 13-16 have been amended; and new claims 25-50 have been added as follows:

- 1           1.       (Amended) A method of determining spatial target probability ~~using~~  
2   ~~a model of multisensory processing by the brain, said method~~ comprising the steps of:  
3               acquiring at least two inputs from a location in a desired environment  
4   ~~where a first target is detected;~~  
5               applying said inputs to a plurality of model units in a map corresponding to  
6   a plurality of locations in said environment;  
7               approximating a posterior probability of ~~said~~ a first target at each of said  
8   model units based on said at least two inputs;  
9               finding a model unit from said plurality of model units with a highest  
10   posterior probability;  
11              choosing a location in said environment corresponding to said model unit  
12   with a said highest posterior probability as a location of a next target.

Claim 5 has been canceled.

1                   6.       (Amended) The method as defined in ~~claim 5~~claim 4, wherein said  
2       posterior probability is approximated using a sigmoid curve function.

1                   7.       (Amended) The method as defined in ~~claim 5~~claim 4, wherein said  
2       posterior probability is approximated using a linear function.

1                   8.       (Amended) The method as defined in ~~claim 5~~claim 4, wherein said  
2       posterior probability is approximated using a bounded linear function.

Claims 9-12 have been canceled.

1                   13.     (Amended) A method of determining spatial target probability using  
2       ~~a neural network model of multisensory processing by the brain~~ a supervised learning  
3       algorithm in a model of a neural network having a plurality of input units, output units  
4       and hidden units connected between said input and output units, said method comprising  
5       the steps of:  
6                   ~~training a plurality model units in a map corresponding to a plurality of~~  
7       ~~locations in a desired environment to output a desired value when an actual target is~~  
8       ~~detected~~ the model neural network to reduce an error between an actual response and a  
9       desired response of the neural network to predetermined inputs from a known location in  
10      a desired environment;

11                    applying ~~at least two inputs from said actual~~ one input associated with a  
12 first target located in said desired environment;  
13                    finding ~~one of said model units~~ an output unit from the plurality of output  
14 units with a highest desired value; and  
15                    choosing a location in said environment corresponding to said ~~model~~ output  
16 unit with said highest desired value as a location of ~~said actual~~ a next target.

1                    14. (Amended) The method as defined in claim 13, wherein said  
2 training step includes:

3                    positioning a training target at a random location in said desired  
4 environment;

5                    ~~acquiring at least two inputs from said training target;~~

6                    ~~applying said at least two inputs said plurality model units in said map and~~

7 ~~obtaining actual responses of said model units~~ at least one input associated with said

8 training target to the model neural network to obtain said actual responses of the model

9 neural network to said training target;

10                    generating said desired responses ~~for said model units~~ of the model neural

11 network to said training target;

12                    finding differences between said actual and desired responses; and

13                    using back-propagation to reduce said differences between said actual and

14 desired responses.

1                   15.    (Amended) ~~A camera~~ An apparatus for automatically tracking a  
2 target in a ~~known~~ desired environment, said system comprising:

3                   at least one ~~audio and at least one video sensors~~ first sensor for receiving  
4 ~~audio and video signals~~ sensory inputs from the target;

5                   a controller ~~for receiving said audio and video signals from said audio and~~  
6 ~~video sensors and determining a probability of the target being at a location, based on~~  
7 said sensory inputs, for locating the target in the environment using a program modeling  
8 ~~mutisensory processing of the~~ a neural network of a brain; and

9                   at least one of ~~a moveable directional audio and video~~ second sensor for  
10 turning to a location in the environment where ~~a the target probability is high as~~  
11 ~~determined~~ has been located by said controller,

12                   wherein said model of said neural network includes a map having a  
13 plurality of model units corresponding to a plurality of locations in the environment for  
14 receiving information from said sensory inputs associated with the target located in the  
15 environment through a plurality of input units and connections between said input units  
16 and said model units.

1                   16.    (Amended) The apparatus as defined in claim 15 wherein said  
2 ~~modeling program approximates a posterior probability of the target given said audio and~~  
3 ~~video signals from the target~~ is located by approximating a posterior probability of the  
4 target given said sensory inputs.

Claims 20-24 have been canceled.

1                   25.   (New) The method as defined in claim 13, wherein the plurality of  
2   output units of the model neural network represent model units in a map corresponding to  
3   a plurality of locations in said desired environment.

1                   26.   (New) The method as defined in claim 14, wherein said step of  
2   using back-propagation includes iteratively adjusting weights associated with the hidden  
3   units.

1                   27.   (New) The method as defined in claim 13, wherein said  
2   predetermined inputs and said at least one input associated with said first target are  
3   sensory inputs.

1                   28.   (New) The method as defined in claim 27, wherein said sensory  
2   inputs include audio and video inputs.

1                   29.   (New) A method of determining spatial target probability using an  
2   unsupervised adaptive algorithm in a model of a neural network, said method comprising  
3   the steps of:

4                   organizing a map into a plurality of model units corresponding to a plurality  
5   of locations in a desired environment for receiving information from sensory inputs

6 associated with a target located in said environment through a plurality of input units and  
7 connections between said input units and said model units;  
8 adjusting said map to increase an amount of said information from said  
9 sensory inputs that are transmitted to said map using an unsupervised learning  
10 mechanism; and  
11 modulating a strength of said sensory inputs associated with said target  
12 based on a correlation between activities of said map and predefined modulation units,  
13 and on anti-correlation between said predefined modulation units and said sensory inputs  
14 associated with said target.

1 30. (New) The method as defined in claim 29, wherein a Kohonen  
2 mechanism is used in said step of adjusting said map.

1 31. (New) The method as defined in claim 30 wherein said Kohonen  
2 mechanism adjusts weights associated with said connections between said input units and  
3 said model units such that each of said model units become specialized for receiving  
4 information indicating a predetermined location in said environment.

1 32. (New) The method as defined in claim 31 wherein said model units  
2 further become specialized for receiving a predetermined modality of said sensory inputs  
3 associated with said target.

1                   33.    (New)  The method as defined in claim 32 wherein said modality  
2   includes at least audio and video inputs.

1                   34.    (New)  The method as defined in claim 32 wherein said modulation  
2   units are predefined according to a modality of said sensory inputs associated with said  
3   target.

1                   35.    (New)  The method as defined in claim 34 wherein said modulation  
2   units modulate said strength of said sensory inputs by multiplying weights associated  
3   with said sensory inputs.

1                   36.    (New)  The method as defined in claim 34 wherein said modality  
2   includes at least audio and video inputs.

1                   37.    (New)  The method as defined in claim 34 wherein at least one of  
2   said modulation units predefined by a first modality of said sensory inputs becomes  
3   active when said map receives information through at least said first modality from said  
4   sensory inputs, said at least one of said modulation units decreases modulation of said  
5   sensory inputs having first modality and increases modulation of said sensory inputs  
6   having modality other than said first modality.



1                   38.   (New) The apparatus as defined in claim 15 wherein said at least  
2   one first sensor includes at least one audio and at least one video sensor.

1                   39.   (New) The apparatus as defined in claim 38 wherein said sensory  
2   inputs are audio and video signals.

1                   40.   (New) The apparatus as defined in claim 15 wherein said at least  
2   one directional second sensor includes at least one of an audio and a video sensor.

1                   41.   (New) The apparatus as defined in claim 15 wherein the target is  
2   located by a supervised learning algorithm in which,

3                   said model neural network is trained to reduce an error between an actual  
4   response and a desired response of said model neural network to predetermined inputs  
5   from a known location in the environment;

6                   sensory inputs associated with the target located in the environment is  
7   applied to said plurality of inputs of said model neural network;

8                   the model units with a highest desired value is found; and

9                   a location in the environment corresponding to said model unit with said  
10   highest desired value is chosen as a location of a next target.

1                   42.   (New) The method as defined in claim 41, wherein said training of  
2   said model neural network includes,  
3                   positioning a training target at a random location in the predefined  
4   environment;  
5                   applying sensory inputs associated with said training target to the model  
6   neural network to obtain said actual responses of the model neural network to said  
7   training target;  
8                   generating said desired responses of the model neural network to said  
9   training target;  
10                  finding differences between said actual and desired responses; and  
11                  using back-propagation to reduce said differences between said actual and  
12   desired responses.

1                   43.   (New) The apparatus as defined in claim 15 wherein the target is  
2   located by an unsupervised adaptive algorithm in which,  
3                   said map is adjusted using a Kohonen mechanism to increase an amount of  
4   information from said sensory inputs that are transmitted to said map; and  
5                   a strength of said sensory inputs associated with the target is modulated  
6   based on a correlation between activities of said map and predefined modulation units,  
7   and on anti-correlation between said predefined modulation units and said sensory inputs  
8   associated with said target.

1                   44.   (New)   The apparatus as defined in claim 43 wherein said Kohonen  
2   mechanism adjusts weights associated with said connections between said input units and  
3   said model units such that each of said model units become specialized for receiving  
4   information indicating a predetermined location in said environment.

1                   45.   (New)   The apparatus as defined in claim 44 wherein said model  
2   units further become specialized for receiving a predetermined modality of said sensory  
3   inputs associated with said target.

1                   46.   (New)   The apparatus as defined in claim 45 wherein said modality  
2   includes at least audio and video inputs.

1                   47.   (New)   The apparatus as defined in claim 45 wherein said  
2   modulation units are predefined according to a modality of said sensory inputs associated  
3   with said target.

1                   48.   (New)   The apparatus as defined in claim 47 wherein said  
2   modulation units modulate said strength of said sensory inputs by multiplying weights  
3   associated with said sensory inputs.

1                   49.   (New) The apparatus as defined in claim 47 wherein said modality  
2   includes at least audio and video inputs.

1                   50.   (New) The apparatus as defined in claim 47 wherein at least one of  
2   said modulation units predefined by a first modality of said sensory inputs becomes  
3   active when said map receives information through at least said first modality from said  
4   sensory inputs, said at least one of said modulation units decreases modulation of said  
5   sensory inputs having first modality and increases modulation of said sensory inputs  
6   having modality other than said first modality.